Evaluating Dual-view Perceptual Issues in Handheld Augmented Reality: Device vs. User Perspective Rendering

Klen Čopič Pucihar
School of Computing and Communications
Lancaster University, UK
k.copicpucihar@lancaster.ac.uk

Paul Coulton
Imagination, Lancaster Institute for the Contemporary Arts
Lancaster University, UK
p.coulton@lancaster.ac.uk

Jason Alexander
School of Computing and Communications
Lancaster University, UK
j.alexander@lancaster.ac.uk

ABSTRACT

In handheld Augmented Reality (AR) the magic-lens paradigm is typically implemented by rendering the video stream captured by the back-facing camera onto the device's screen. Unfortunately, such implementations show the real world from the device's perspective rather than the user's perspective. This dual-perspective results in misaligned and incorrectly scaled imagery, a predominate cause for the dual-view problem with potential to distort user's spatial perception. This paper presents a user study that analyzes users' expectations, spatial-perception, and their ability to deal with the dual-view problem, by comparing device-perspective and fixed Point-of-View (POV) user-perspective rendering. The results confirm the existence of the dual-view perceptual issue and that the majority of participants expect user-perspective rendering irrespective of their previous AR experience. Participants also demonstrated significantly better spatial perception and preference of the user-perspective view.

Categories and Subject Descriptors

H5.1. Information interfaces and presentation (e.g., HCI): Artificial, augmented, and virtual realities.

Keywords

Perception; perceptual issue; user expectation; user-perspective; device-perspective; dual-perspective; dual-view; user study; AR; handheld; mobile; spatial-perception; rendering;

1. INTRODUCTION

Handheld Augmented Reality (AR) is normally presented using the magic-lens paradigm [5], in which a transparent lens reveals an enhanced segment of the real-world. The widespread adoption and increased processing power of handheld devices means such form of AR is readily available via simple application downloads. AR renderings in such implementations generally utilize the device's single back-facing camera and screen.

However, the typical virtual transparency provided by the device's camera and screen, known as device-perspective rendering, does not match the clear glass-pane transparency suggested by the magic-lens paradigm. This is due to: the device's non-centered camera, differences in the Field-of-View (FOV) between the device and the user, differences in the angular offset between the device and the user, and the monocular camera capture and rendering. The primary issue introduced by this AR setup is the dual-view problem.

The dual-view problem mainly occurs because of the dual-perspective—the device's camera captures the real world scene from a different perspective to the observer's. The dual-perspective may result in imagery that is misaligned and/or incorrectly scaled. This effect is demonstrated in Figure 1a that shows device-perspective magic-lens—imagery that does not match with the surrounding visual information—and Figure 1b that shows user-perspective magic-lens—imagery that matches what the user would see if the device acted as a clear glass pane. Such dual-view effect is expected to distort users' spatial perception.

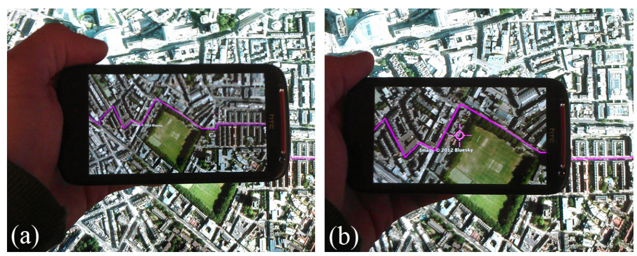


Figure 1: (a) The dual-view problem caused by device-perspective magic-lens rendering. (b) User-perspective magic-lens rendering where such dual-view problem is not present.

User expectations, formed through both 'hands-on' experience and 'hands-off' media exposure, can also influence the severity of distortion in users' spatial perception. Typical camera-based activities on mobile devices (taking photos, shooting videos, scanning barcodes) use device-perspective rendering, as do, to the best of our knowledge, all commercial AR systems. Interestingly though, the majority of advertising for AR systems present the (possibly misleading) user-perspective imagery, creating confusion between the perceived affordances [20]. This mismatch of expectations leads to a number of interesting questions: Does the dual-view problem coupled with users' expectations result in a dual-view perceptual issue? Are the expectations of new and returning AR users different? How successful are users in dealing with perceptual issues that may arise?

This paper aims to address these questions through two user studies that compare device-perspective and fixed Point-of-View (POV) user-perspective rendering on a commercially available handheld device. The first study explores user expectations to confirm the existence of the dual-view perceptual issue. The second study investigates spatial perception by asking users to relate augmented content to the real environment.

ICMI '13, December 09 - 13 2013, Sydney, NSW, Australia

ACM 978-1-4503-2129-7/13/12…$15.00.
http://dx.doi.org/10.1145/2522848.2522885

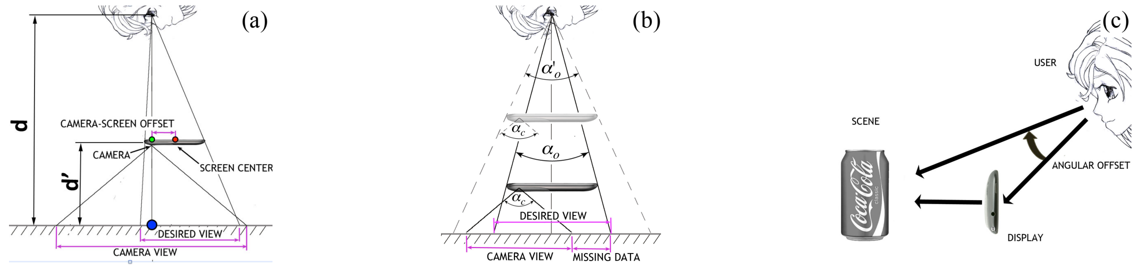


Figure 2: (a) The effect of non-centered scene capture: the blue dot is rendered at the position of the red dot. (b) The difference between the static FOV of the phone's camera (α_c) and the dynamic FOV the observer would expect if the phone acted as a transparent glass pane (α_o to α'_o) (c) Angular offset, a result of holding the phone non-perpendicularly to the observer's POV.

Our results show: (1) The existence of the dual-view perceptual issue; (2) The dual-view problem distorts users' spatial perception; (3) Users' expectations do not significantly affect their ability to deal with distorted spatial perception; (4) Even after participants identified their expectations were wrong, they continue to demonstrate a slow learning process when dealing with the dual view perceptual issue; and (5) User-perspective rendering is subjectively preferred.

2. PROBLEM ANALYSIS

Virtual transparency in handheld AR is typically implemented by rendering the video stream captured by the back-facing camera onto the device's screen. Such rendering does not match the ("ideal") clear glass pane transparency suggested by the magic-lens paradigm. If ideal transparency is not achieved, this may result in the dual-view problem. The differences between ideal and implemented transparency are mainly caused by: (1) The monocular camera scene capture and rendering introducing the depth perception problem, and (2) The camera capturing the real world scene from a different perspective than the observer's, causing the dual-perspective problem.

2.1 The Depth Perception Problem

Monocular scene capture and rendering distorts *binocular disparity*[1]. As a result, the user's ability to estimate depth may be affected [7]. To cope with binocular disparity, users typically employ more important depth cues such as motion parallax and depth ordering [1, 8, 14] to reconstruct scene depth. As other, more important, depth cues are available we expect binocular disparity to effect the user's spatial perception less than the dual-view problem. Hence, the remainder of the paper focuses on addressing this issue.

2.2 The Dual-View Problem

The dual-view problem is predominantly a result of perspective differences between the camera and the observer, a result of typical handheld AR setups as described below:

1. *Non-centered Scene Capture:* the camera is physically positioned off-center from the phone screen, with the video stream still presented in the center of the display. This can be observed in Figure 2a, where the blue dot in the scene is rendered at the position of the red dot on the display.

2. *Differences in the Field of View (FOV)*: the phone camera's FOV (α_c on Figure 2b) is different to the FOV the observer would see if the magic lens acted as a transparent glass pane (α_o on Figure 2b). Additionally, and contrary to the dynamic FOV of the transparent glass pane, the camera has a static FOV (in Figure 2b, the transparent glass pane FOV changes with magic lens distance from α_o to α'_o).

3. *Angular Offset of Views:* when the device is held at a non-perpendicular angle to the observer's Point-of-View (POV, see Figure 2c).

The cumulative causes of the dual-perspective may result in imagery that is misaligned and/or incorrectly scaled (see Figure 1a), when compared to what the user would see if the device acted as a clear glass pane. These differences are expected to affect users' spatial perception. This in turn is expected to reduce their ability to relate augmented content to the real-world scene (observe how the phone screen fits to the surroundings in Figure 1a and 1b). Beside dual-perspective, other factors, such as different focal and disparity planes of augmented and real view may also contribute to dual-view problem [14], however within the context of this study the magnitude of these effects is expected to be much smaller. Hence, the remainder of the paper focuses on the dual-view problem caused by dual-perspective.

3. BACKGROUND

Dual-view and depth perceptual issues are well known within the AR community [14], however the number of studies examining the effects such issues have on user performance and satisfaction are limited. In the case of the dual-view perceptual issue, the majority of previous work is limited to theoretical analyses and user experiments within Mixed Reality (MR) simulators [3, 21]. However, MR simulators take users out of their real environment, potentially undermining previous real world experiences, which amongst other influences, shape users' perceived affordances [20].

Understanding users' expectations and their ability to deal with perceptual problems is vital in order to improve the usability of handheld AR interfaces especially given Olsson's et al. report that AR interfaces are inconsistent and questionable for their pragmatic usefulness in everyday life [22].

3.1 Tracking and 3D Scene Reconstruction

One of the key elements of any handheld AR system is robust six Degrees of Freedom (DOF) camera pose tracking, nowadays possible using feature descriptor-based approaches, where natural features from predefined environments can be used for camera pose tracking [26]. Other approaches take advantage of the improved sensing capabilities of mobile devices to merge sensory and visual data to support camera pose initialization and tracking [6, 15, 16]. In addition to these offline systems that use pre-defined environmental maps, there are online tracking solutions known as Simultaneous Localization and Mapping (SLAM).

In SLAM systems, the 3D scene is reconstructed on the fly. Even though successful demonstrations of SLAM have been presented on commercial handheld devices [13] the difficulty and

1 The difference in image location of the object due to left and right eye parallax

computational cost of 3D reconstruction continues to limit the complexity of the generated 3D maps. Additionally, such maps are built incrementally through multi-frame tracking and thus are not able to cope with fast changing environments. In the case of user-perspective rendering, any elements that are not part of the mapped 3D scene (i.e. a hand interacting with the AR workspace) would be rendered incorrectly. Encouragingly, Newcombe et al. [19] and Izadi et al. [12] demonstrate that dense model reconstructions of dynamic environments is possible using Kinect sensors, should depth cameras become available in handheld devices.

3.2 User-perspective Rendering Prototypes

Recently, several user-perspective rendering prototypes have emerged within the research community [3, 11]. Baričević et al. implemented user-perspective rendering via per-pixel 3D scene reconstruction using a Kinect depth sensor and face-tracking using a Wiimote [3]. Their solution provides a real-time per-pixel 3D scene, unfortunately, limitations of the Kinect depth sensor range prevent correct rendering of the user's hand while interacting with the AR scene. Until depth sensors become available on handheld devices this method is unsuitable for deployment.

Hill et al. implemented user-perspective rendering using a large FOV back facing camera from which rectilinear wide FOV images are created. Based on the viewer's POV (estimated by the front facing camera) an appropriate section of the large FOV image is selected [11]. Such an approach creates imagery with different perspective distortion to the one seen by the observer and allows correct alignment of the observer's frustum at a single distance. Consequently, correct rendering of the hand interacting within the AR scene is not possible. Even though a front facing camera is commonly available on handheld devices, and previous studies of head pose estimation using a single camera show it is achievable at reasonable levels of accuracy [2], such tracking adds a considerable computational cost to the already resource limited camera pose tracking and mapping tasks. Hence, such solutions have yet to be proven on a single handheld device.

3.3 User Studies

User studies of handheld AR using physical devices generally compare device-perspective magic-lenses to alternative non-AR interfaces [10, 17, 24, 25].

Rohs et al. [24] demonstrated that performance in large image navigation (such as maps) on a mobile phone is higher in physical movement interfaces (peephole and AR interfaces) when compared to virtual movement interfaces (joystick interface). Perhaps surprisingly, they reported no significant difference between the dynamic peephole [18] map navigation, where the phone surroundings exhibited a random pattern, and a device-perspective magic-lens interaction where phone surroundings exhibited meaningful information (i.e. a paper map). Rohs et al. conclude that *"switching between the two layers of visual presentation incur higher costs than expected as with each switch of layers the user's eyes have to refocus on the new depth and locate the intended object on the new layer"* [24]. Even though in later research on item density effects [25] the surrounding visual context was found to be of significant importance, it is clear that relating augmented content to the real world in handheld AR setups is a non-trivial task. Nevertheless, Rohs et al. did not consider the dual-view problem, a result of dual-perspective, as a factor contributing to such an outcome.

Maintaining a sense of surrounding context is also important in the context of document navigation [4, 23]. Baudisch et al. identified the ability to seamlessly merge context with detail views, a common feature in focus+context screens, as something that significantly improves the usability of such systems [4]. This supports the notion that AR interfaces should seamlessly merge a magic-lens with surrounding context.

Baričević et al. [3] compared user and device-perspective magic-lens interfaces using selection and search tasks in an MR simulator. In this instance the user's hand was represented by a dot, potentially reducing the effects of the dual-view problem. The study showed participants' performed selection tasks on tablet-sized displays significantly faster using user-perspective rendering.

4. FIXED POINT-OF-VIEW USER-PERSPECTIVE RENDERING

Fixed point-of-view (POV) user-perspective rendering is a technique that produces user-perspective rendering without the need for head pose tracking. Fixed-POV user-perspective rendering was employed in this research as at the time of running the study, full user-perspective rendering was not yet realizable on handheld devices. This rendering is used in the comparative studies in the remainder of this paper, and is described here.

To eliminate the need for head pose tracking, the observer's POV is fixed above the physical center of the magic-lens screen. This allows the rendering software to assume the observer's POV, but means the observer will only see the correct imagery when their head is in the correct position (see Figure 3). As a result, removing the need for head pose tracking simplifies the implementation of user-perspective rendering system which only requires six DOF camera pose tracking and a 3D model of the scene.

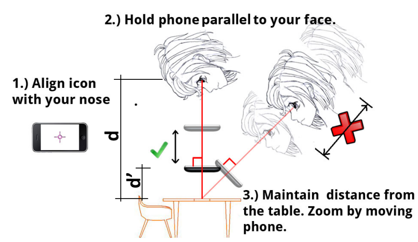


Figure 3: Instructions page summarizing fixed-POV user-perspective rendering assumptions.

Fixing the observer's POV constrains the user to holding the phone at an angle perpendicular to the observer's POV (see Figure 3). As a result, angular offset does not contribute to the dual-view problem. However, this solution assumes that holding the phone at a perpendicular angle to the observer's POV is the most intuitive interaction scenario when using handheld AR. Additionally, in the case of tabletop sized environments, distances between the magic-lens and the interactive workspace (on Figure 2a and 3 denoted as *d'*) are expected to remain small, reducing the effect of angular offset on the rendered imagery To avoid limiting all six DOF between the observer's POV and the device, a constant distance between the observer and the AR workspace is assumed (*d* on Figure 2a and Figure 3) and needs to be initialized at system startup. Initialization can be made by moving the phone up to observers POV whilst camera pose is tracked. This design then allows the user to zoom by moving the phone up and down, changing the distance between the lens and the tabletop surface (*d'* on Figure 2a and 3).

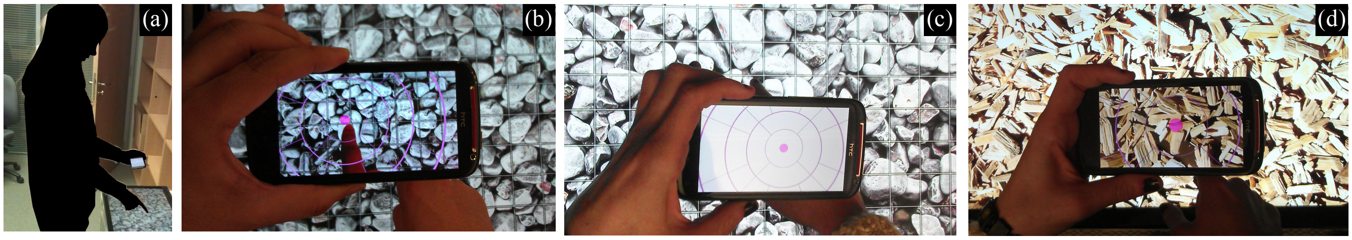


Figure 4: (a) Experimental setup: participant holding mobile phone in their left hand and touching the interactive surface with their right hand; (b) Training phase task: device-perspective magic-lens render where the hand is visible. The effect of the dual-view problem on the hand is obvious (misalignment between index finger and the hand); (c) User Study A task, with all visual cues removed from magic-lens render; (d) User Study B task, only the hand is not rendered within the magic-lens view.

The fixed-POV user-perspective rendering system employs incremental 3D scene reconstruction or predefined 3D models, as real-time per-pixel 3D reconstruction is not yet realizable on handheld devices. Planar scene models are created on the fly by rectifying camera-captured images or by using predefined textures (i.e. a digital version of the printed media). Irrespective of the method used, such maps are limited to planar environments and are not updated with every captured frame. Consequently, such a system cannot render correct user-perspective imagery of a user's hand interacting with the magic-lens view.

5. USER STUDY A: UNDERSTANDING EXPECTATIONS

The first of our two user studies aims to determine users' current spatial expectations of handheld AR and confirm the existence of the dual-view perceptual issue, a result of differences between user's expectations and system behavior. We also wish to investigate whether users' expectations affect the severity of spatial distortion caused by the dual-view problem, a result of dual-perspective, and determine if this distortion can be overcome by introducing fixed-POV user-perspective magic-lenses.

We predict: (1) Users with previous experience of handheld AR will expect device-perspective rendering, whereas users with no previous experience will expect user-perspective rendering; (2) User expectations will influence their ability to deal with distorted spatial perception; and (3) Users will demonstrate better spatial perception when interacting with fixed-POV user-perspective magic-lens.

5.1 Experimental Setup and Design

To explore users' spatial expectations in handheld AR, we asked participants to touch the real-world location of an augmented target. Participants revealed the location of the target using AR software on mobile device, with finger taps recorded on an interactive surface. We ran a within-subjects experiment, changing only a single variable, the type of magic-lens. This had one of two values:

- *Device-perspective magic-lens:* the lens shows the view captured from the handheld device's camera.
- *Fixed-POV user-perspective magic-lens*: as described in section 4.

The physical setup (see Figure 4a) consisted of a 24-inch horizontal interactive surface mounted at table height and a handheld mobile device. The interactive surface is the AR workspace, where different surface textures are presented to the user and touch events are recorded. The magic-lens is rendered on an off-the-shelf HTC Sensation mobile phone with a 4.3 inch screen with 16:9 aspect ratio. Throughout the experiment users stand and hold the phone in landscape orientation.

5.2 Experimental Task

Each task consists of participants finding and tapping on an augmented-reality target. The target is visible only on the magic-lens and is randomly placed on the interactive surface for each task repetition. Targets have a diameter of 1.2 cm (Figure 4c). A task begins when the user taps the phone screen (enabling target discovery) and ends once the user indicates the target position on the interactive surface. We specifically choose to remove all visual links between the real-world and the magic-lens rendering (compare Figure 4b and 4c), both the background scene and the user's hand, as this would immediately bias expectations. We also choose to use touch to indicate the physical location of augmented objects as we believe this to be the most intuitive method for demonstrating the relationship. Positional error, the distance from the center of the target to the center of touch event, is recorded for each task. The rendered target position on the phone at the time of the touch event is also recorded.

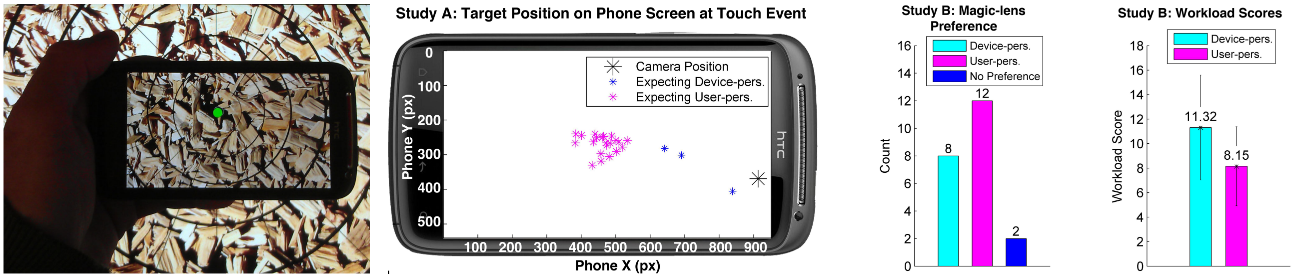
5.3 Experimental Procedure

To begin, participants completed a setup and training period. The distance between the observer's POV and the interactive surface is measured (*d* in Figure 3). This distance is required to initialize the fixed-POV user-perspective rendering. Participants are instructed to keep the phone perpendicular to their POV throughout the experiment. This decision was made to keep the experimental setup conditions unchanged throughout the study even though this is only required when interacting with fixed-POV user-perspective magic-lens. To familiarize participants with handheld AR, participants performed three target selections using device-perspective rendering. During training the participant's hand is visible to allow adequate task explanation (see Figure 4b). Participants then move to the real experimental tasks.

Participants always started with the device-perspective magic-lens. They were instructed to focus only on accuracy (and were under no time pressure) and so we only recorded positional error. No feedback regarding participants' success was given in order to avoid influencing their initial expectations of handheld AR. After performing seven task repetitions with device-perspective rendering, they were asked to explain the strategies they used for mapping the augmented target from the phone screen to the interactive surface. Before moving to their seven tasks with fixed-POV user-perspective rendering, participants were shown the correct use of this rendering method (see Figure 3).

5.4 Participants

The study was conducted with 24 participants, 8 female and 16 male, aged between 21 and 45. Of the 24 participants, 17 indicated they already understood the concept of AR and 14 stated they had previous experience using handheld AR applications.



(a) (b) (c) (d)

Figure 5: (a) User Study B task feedback shown after each task attempt: the augmented target is rendered in black with the green dot showing the position of the touch event; (b) User Study A, device-perspective magic-lens results: Average target position on phone at touch event; (c) User Study B results: User preferences. (d) User Study B results: Average NASA TLX workload scores;

5.5 Results

Participants successfully completed all 336 tasks (24 participants × 2 renderings × 7 repetitions). Overall, the study results partially confirmed our predictions. However, due to only a small number of subjects expecting device-perspective rendering, drawing final conclusions regarding the effect of users' expectations on distorted spatial perception is difficult. Irrespective of users' expectations, participants demonstrated better spatial perception when interacting with the fixed-POV user-perspective magic-lens. More detailed results follow.

5.5.1 Rendering Expectations

Using the task completion strategies reported within the study questionnaire, the prediction that users expect user-perspective rendering was partially confirmed. All 10 participants with no previous AR experience expected user-perceptive rendering. However, contrary to expectations, of the 14 participants who had previous AR experience, the majority ~78% (11) (95% confidence interval [0.571, 1.000]) expected user-perspective rendering.

The same trend is confirmed when analyzing the rendering position of the target on the phone's screen. Figure 5b shows the position on the phone screen of the target at the time of the touch event. In all cases where task completion strategies provided clues of device-perspective rendering expectations, participants identified camera position as not positioned in the screen center (blue data points on Figure 5b), whereas participants expecting user-perspective rendering expected the camera to be located in the center of the screen (pink data points on Figure 5b).

5.5.2 Augmented-to-Real-World Mapping Accuracy

Users expecting device-perspective rendering identified that the camera was not in the center of the screen and tried to accommodate for the offset when interacting with device-perspective magic-lens. Even though they achieved slightly better accuracy (User Study A on Figure 6c), a one-way unbalanced ANOVA does not detect a significant difference in accuracy between the group expecting device-perspective and user-perspective rendering ($F=0.018$, $P=0.896$). Even though the trends show no significance between the two groups, using only 3 subjects as a sample of the population expecting device-perspective rendering limits our ability to draw final conclusions regarding how subjects' expectations influence their ability to deal with distorted spatial perception. Nevertheless, irrespective of user's expectations, a paired sample T-test confirmed participants demonstrated significantly better spatial perception when interacting with fixed-POV user-perspective magic-lens ($P<.001$, 95% CI [1.82, 2.73] cm) (see Study A on Figure 6a).

6. USER STUDY B: SPATIAL PERCEPTION

User study A established a general expectation of, and increased accuracy with, user-perspective rendering. In this study we wished to more deeply explore the effect of the dual-view problem, a result of dual-perspective situation, on users' ability to relate augmented content to the real world. In the previous study, all visual links between the magic lens and the interactive surface were removed. This study returns these visual links to explore the influence of surrounding visual context on spatial perception using the two types of rendering.

Our predications are: (1) Users will demonstrate better spatial-perception when interacting with the fixed-POV user perspective rendering, (2) Subjects will be successful at learning how to deal with the dual-view problem and will improve their accuracy with practice, (3) As better spatial-perception is expected in the fixed-POV user-perspective rendering, the perceived task workload scores of NASA TLX questionnaires are expected to be smaller than when using device-perspective rendering, (4) Participants are expected to prefer the fixed-POV user-perspective magic-lens.

6.1 Experimental Setup and Design

To explore users' spatial understanding, participants were asked to relate augmented content to the real world. We used an identical experimental setup and design as User Study A, with the differences in magic-lens rendering noted in the following section.

6.2 Experimental Tasks

As per User Study A, participants were asked to touch the position of an augmented target on an interactive surface. In this study, the visual links between the magic lens and the surrounding context were re-established. In study A, the magic lens rendered a blank screen, here we render the scene below the lens (compare Figures 4c and 4d). The tasks otherwise remain identical to User Study A.

In addition to positional error, we also measured participants' performance via task completion time (the time between the appearance of the target on the phone screen and the touch on the interactive surface). To isolate the effect of the dual-view problem on users' spatial understanding, we continue to not render the, user's hand. As a result of removing the hand, relating augmented content to the real world is limited to two strategies namely: using spatial perception to predict augmented content location, and by image comparison where visual links between the real world and the phone screen are used to identify the augmented target position. We used complex image patterns (see Figure 5a) and set a minimum accuracy threshold (defined by a pilot study to be 1.2 cm) to ensure appropriate task difficulty and prevent participants from using a single solve strategy.

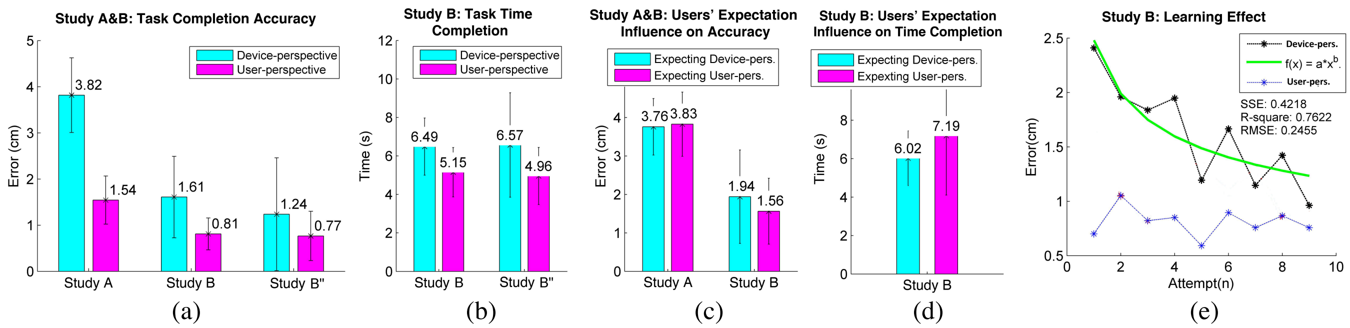


Figure 6: (a) Accuracy for device and user-perspective magic-lens. B'' shows the final three attempts; (b) Task completion time; (c) Accuracy and (d) task completion time based on participant expectations from device-perspective magic-lens interaction; (e) Learning effect: Average error in cm for each attempt of Study B with a power-curve mapped to user-perspective data points.

6.3 Experimental Procedure

As per User Study A, participants first completed a setup and training period. To begin, participants are given a short demonstration illustrating the differences between the two types of magic-lens. Participants are then given a maximum of 2 minutes to familiarize themselves with each of the two magic-lens renderings by interacting with a map (see Figure 1).

Following training, the selection task is demonstrated to the user. In this study, we provide the participant with feedback on the accuracy of their selection to enable a learning process and to ensure they continue to make as accurate selections as possible. The results are shown to participants by displaying the augmented target (black rendering on Figure 5a) and the touch event (green dot on Figure 5a). Participants were asked to be as quick and as accurate as possible with minimal acceptable accuracy of 1.2 cm.

Participants made nine task repetitions with each magic-lens type. Contrary to study A, here the order in which magic-lenses are tested is randomized. For every three task repetitions the surface changes in order to minimize the participants' ability to learn the pattern. The same surfaces are used for both magic-lens types and are presented in the same sequence to all participants. After completing all tasks, participants are asked to select their preferred magic-lens type and complete a NASA TLX [9] workload assessment questionnaire.

6.4 Participants

The same participant group and physical experimental setup was used here as employed in User Study A. Participants always performed User Study A before moving on to User Study B.

6.5 Results

Participants successfully completed all 432 tasks (24 participants × 2 renderings × 9 repetitions).

6.5.1 Augmented-to-Real-World Mapping Accuracy

The results of the analysis confirm the majority of our predictions. Running a paired sample T-test confirms participants were significantly more accurate (P<.0.001, 95% CI[0.45, 1.14] cm]) and significantly faster (P<.001, 95% CI[0.85, 1.8] seconds) when completing the task using the fixed-POV user-perspective magic lens (Study B on Figure 6a and Figure 6b).

6.5.2 Learning Effects

The results also confirm that learning only occurs in device-perspective rendering where the dual-view perceptual problem exists (see Figure 6e). Fitting a power function revealed an overall R-square fit of 0.7522. Even though there is no obvious learning in user-perspective magic-lens, after 9 attempts the accuracy achieved in device-perspective view does not surpass the accuracy achieved when interacting with user-perspective view (Figure 6e).

Even though the learning curve does not level off completely by the 6th observation of device-perspective magic-lens use, the last three observations are used to estimate participants' performance after the learning phase is completed. Even though participants continue to be more accurate on average in fixed-POV user-perspective magic-lens (Study B" on Figure 6a), the paired sample T-test failed to detect significance (P=.064, 95% CI [-0.023, 0.98]cm). However, a significant reduction in task completion time (Study B" on Figure 6b), continued to be detected when interacting with fixed-POV user-perspective magic-lens (P=.028, 95% CI [0.2, 3.0] s).

6.5.3 Rendering Expectations

The influence of participants' expectations on their ability to deal with distorted spatial perception is analyzed by comparing the performance of the two groups. The same groups are used as in Study A. The results show that users expecting device-perspective rendering demonstrated slightly lower accuracy (Study B in Figure 6c) and faster task completion (Study B in Figure 6d). However, running a one-way unbalanced ANOVA does not detect a significant difference in time (F=0.404, P=0.532) and accuracy (F=0.466, P=0.5) between the group expecting device-perspective and user-perspective rendering. However, again, using only 3 subjects as a sample of the population expecting device-perspective rendering limits our ability to draw final conclusions.

6.5.4 Preferences

Analyzing the questionnaires revealed that participants preferred the fixed-POV user-perspective magic lens (see Figure 5c), however, running Wilcoxon rank test did not show a significant difference in users' preference (P=0.1). However, when looking at perceived workload scores (see Figure 5d), paired sample T-test statistics showed there is significant decrease in perceived workload scores when interacting with the fixed-POV user-perspective magic lens (P<0.001, 95% CI [1.8, 2.7]).

7. DISCUSSION

7.1 Understanding Rendering Expectations

The results show that users interacting with handheld AR expect the device to act as a transparent glass-pane. Previous experiences of camera-based activities on mobile devices (e.g. taking photos) that typically show device-perspective rendering, did not change this expectation.

By analyzing task completion strategies we classified participants into two groups: (1) the group expecting device-perspective rendering, and (2) the group expecting user-perspective rendering.

The majority of new and returning handheld AR users demonstrated user-perspective rendering expectations. Further, when participants who expected device-perspective rendering interacted in this mode, they did not outperform participants who expected user-perspective rendering. This confirms that the dual-view perceptual issue is present in both groups.

When interacting with the device-perspective magic-lens, the group expecting user-perspective rendering made the mistake of assuming an incorrect camera position. Coupling strategy analysis and observing Figure 5b shows that participants aligned the dot in the device's screen center and touched below the phone center, therefore assuming the incorrect camera position (see Figure 2a).

The group of participants that named strategies linked to device-perspective rendering expectations did not demonstrate a correct understanding of the dual-view problem introduced by this rendering. This group correctly assumed the camera position, but wrongly assumed the camera imagery was rendered at the camera position and not center of the device's screen (see Figure 2a). Coupling strategy analysis and observing Figure 5b shows that participants expecting device-perspective rendering aligned the target towards the camera position, thus the correct camera position was assumed. By not outperforming those expecting the user-perspective magic-lens, it is clear that the target was not expected below the center of the device's screen, thus they did not successful account for the non-centered camera position (explained in Figure 2a).

7.2 Augmented-to-Real-World Mapping

The results show that fixed-POV user-perspective rendering can be successfully used. In both studies, users' spatial perception significantly improved when using user-perspective rendering over device-perspective rendering. This confirms the results of Baričević et al.'s MR simulator-based study that reported user-perspective rendering improves spatial perception [3].

Learning was only expected in device-perspective rendering as this is the only mode where dual-view perceptual issues are present. The second study confirmed our expectations; learning was only obvious when participants interacted with the device-perspective magic-lens. The speed of learning with the device-perspective magic-lens is slower than expected (see Figure 6e), as users are continuing to improve even after six task repetitions. The improved task performance demonstrates that users learn how to deal with dual-view problem. Overall though, participants continue to perform better when interacting with the fixed-POV user-perspective magic-lens.

7.3 Study Limitations

We used fixed-POV user-perspective rendering to evaluate user-perspective magic-lenses. Here, the observer's point of view is fixed at a predefined position, constraining the user to hold the phone perpendicular to the real-world environment (see Figure 3). Until full user-perspective rendering is feasible on handheld devices, we believe this to be the best method of evaluating the dual-view problem.

Real-time per-pixel 3D mapping is not currently possible on handheld devices, limiting user-perspective rendering solutions to static environments. This prevents the correct rendering of a hand interacting within an AR workspace. Removing the hand from the magic-lens render is an artificial constraint in tabletop environments where one hand would typically be available for interaction. Irrespective of this limitation we believe rendering a representation of the hand would produce similar results although this needs to be verified in the future. Additionally, there are many instances of AR where the workspaces are out-of-reach for interaction. In our instance, removing the user's hands avoided bias when testing user expectations.

Finally, even though qualitative and quantitative results show that fixed-POV user-perspective rendering was successfully adopted, these results were obtained within a controlled environment. Successful adoption of fixed-POV user-perspective rendering in real world settings remains to be confirmed.

7.4 Implications for Research

The two studies demonstrate that participants were unable to deal with dual-view problem introduced by the device-perspective magic-lens. Further, even after participants identified that their initial expectations were wrong they took considerable time to correct their input actions. The two studies therefore reveal that users have a hard time dealing with dual-view perceptual issue, particularly with the effect caused by the non-centered camera (explained in Figure 2a).

This discovery is important for the handheld AR research community as it uncovers the importance of the camera position on the handheld device. Repositioning the camera to the screen center would solve this problem. However, device designers have strong reasons the current camera position (likely based around manufacturing processes). This opens up the question of designing the optimal visualization to help users deal with the non-centered camera. Such visualizations will become increasing important as the lower complexity of device-perspective rendering solutions and the likely persistence of non-centered cameras, mean the dual-view perceptual issue is here to stay.

These results are also important in the context of AR systems with fixed displays and moving observers. In such AR settings the display acts as a mirror showing a reflection of the real world augmented with digital content (e.g. Sony's Playroom system). These systems are commonly implemented using a single static camera positioned at an offset from the TV screen center; hence producing device-perspective-reflection, which is different to what the user would see if looking into a mirror. The mismatch between the user's expectations and system's is likely to reoccur.

7.5 Implications for Users

This work has uncovered users' expectations and their ability to deal with the dual-view perceptual issue. We hope this will encourage the redesign of interaction with AR interfaces. Specifically this work: (1) Promotes the development of novel hybrid AR interfaces, enabling different types of magic-lenses; (2) Encourages the design of novel visualization methods to help participants deal with the camera-scene offset; and (3) Petitions for a change in the hardware design of mobile handsets, by placing the camera lens into the center of the device screen, reducing perceptual problems in handheld AR.

8. CONCLUSION AND FUTURE WORK

This paper studied the dual-view problem, a result of dual-perspective situation, in detail by analyzing users' expectations and comparing device- vs. fixed-POV user-perspective rendering on a commercially available mobile device.

The initial assumption of user-perspective rendering being the most intuitive magic-lens view was confirmed, verifying the dual-view perceptual issue. The 24 participants provided substantial qualitative and quantitative evidence to support this assumption. Contrary to our initial expectations, the majority of experienced and non-experienced AR users expected user-perspective rendering. All participants expecting device-perspective rendering had previous experience with handheld AR, however, they did not succeed in outperforming other participants when interacting with the device-perspective magic-lens render.

Throughout the study users demonstrated successful adoption of the fixed-POV user-perspective rendering. Participants favored user-perspective over device-perspective rendering and demonstrated significantly better spatial perception in this mode. As predicted, significant learning only occurs in device-perspective rendering, however, even after participants identified that their initial expectations do not match the system, the learning process remained slow.

Besides distorting the user's spatial perception, the dual-view problem may also affect the use of the surrounding visual context (area around the perimeter of the magic-lens). This is important because effective use of the surroundings is expected to facilitate a more fluid interaction which may affect usability of handheld AR interfaces. Future research should also look at designing hybrid AR interaction systems where different types of magic-lens would combine to achieve overall better system usability. Additionally, the studies revealed that participants struggled to deal with the non-centered camera, thus visualizations representing the camera-screen offset should be designed and tested. As the dual-view perceptual issue detected here is also likely in AR settings with a fixed display and a moving observer (i.e. Sony Playroom system), it should be verified in future user studies. Finally, advances in user-perspective rendering solutions on handheld devices should use the front facing camera to support dynamic POV user-perspective rendering as well as employ hand tracking algorithms to correctly reintroduce the user's hand into the AR workspace.

9. Acknowledgments

Authors would like to thank Slovene Human Resources Development and Scholarship Fund for PhD funding that made carrying this work out possible.

10. REFERENCES

[1] Arthur, K. W., Booth, K. S., and Ware, C. "Evaluating 3D task performance for fish tank virtual worlds," ACM Transactions on Information Systems, vol. 11, no. 3, pp. 239–265, 1993.

[2] Asteriadis, S., Karpouzis, K., and Kollias, S., "Head pose estimation with one camera, in uncalibrated environments," In EGIHMI '10, pp. 55–62.

[3] Baričević, D., Lee, C., Tobias, H., and Bowman, D.A., "A Hand-Held AR Magic Lens with User-Perspective Rendering," In ISMAR'12, pp. 197 - 206.

[4] Baudisch, P., Good, N., Bellotti, V., and Schraedley, P., "Keeping Things in Context : A Comparative Evaluation of Focus Plus Context Screens, Overviews , and Zooming," In SIGCHI '02, 2002, pp. 259–266.

[5] Bier, E.A., Stone, M.C., Pier, K., Buxton, W., and DeRose, T. D., "Toolglass and magic lens: the see- through interface," In SIGGRAPH '93, pp. 73–80.

[6] Čopič, P., K., Coulton, P. and Hutchinson, D., Utilizing Sensor Fusion in Markerless Mobile Augmented Reality. In MobileHCI '11, pp. 663-666.

[7] Čopič, P., K., Coulton, P., and Alexander, J., "Creating a Stereoscopic Magic-lens to Improve Depth Perception in Handheld Augmented Reality". In MobileHCI '13, pp. 448-451.

[8] Cutting, J., "Reconceiving Perceptual Space, in Perceiving Pictures: An Interdisciplinary Approach to Pictorial Space", MIT Press: Cambridge, pp. 215-238, 2003.

[9] Hart, S. and Staveland, L., "Development of NASA- TLX (Task Load Index): Results of empirical and theoretical research," Human Mental Workload, 1, 1988.

[10] Henze, N., Boll, S., "Who's That Girl? Handheld Augmented Reality for Printed Photo Books", In INTERACT '11, pp. 134-151.

[11] Hill, A., Wilson, J., Davidson, B., Gandy, M., and Macintyre, B., "Virtual Transparency: Introducing Parallax View into Video See-through AR," In ISMAR '11, pp. 239–240.

[12] Izadi, S., Kim, D., and Hilliges, O., "KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera," In UIST '11, pp. 559--568.

[13] Klein, G. and Murray, D., "Parallel tracking and mapping on a camera phone," In ISMAR '09, pp. 83–86.

[14] Kruijff, E. and Ii, J. E. S., "Perceptual Issues in Augmented Reality Revisited," In ISMAR '09, pp. 3–12.

[15] Kurz, D. and Benhimane, S., "Gravity-Aware Handheld Augmented Reality," In ISMAR '11, pp. 111–120.

[16] Lee, W., Park, Y., Lepetit, V., and Woo, W., "Point-and-shoot for ubiquitous tagging on mobile phones," In ISMAR 10', pp. 57–64.

[17] Liao, C., Liu, Q., 13Liew, B., and Wilcox, L., "Pacer: fine-grained interactive paper via camera-touch hybrid gestures on a cell phone," In CHI '10, pp. 2441–2450.

[18] Mehra, S., Werkhoven, P., and Worring, M., "Navigating on Handheld Displays: Dynamic versus Static Peephole Navigation," In TOCHI '06, pp. 448– 457.

[19] Newcombe, R., Lovegrove, S., and Davison, A., "DTAM: Dense tracking and mapping in real-time," In ICCV '11, pp. 2320–2327.

[20] Norman, D. A. The Design of Everyday Things. Published by Basic Books, September 17, 2002. ISBN- 10: 0465067107.

[21] Oh, J. and Hua, H., "User evaluations on form factors of tangible magic lens," In ISMAR '06, pp. 23–32.

[22] Olsson T., and Salo M., "Online user survey on current mobile augmented reality applications". In ISMAR '11, pp. 75–84.

[23] Robbins, D. C., Cutrell E., Sarin R., and Horvitz E., "ZoneZoom: map navigation for smartphones with recursive view segmentation," In AVI '04, 2004, pp. 231–234.

[24] Rohs, M., Schöning, M., Raubal, G., Essl, and A. Krüger, "Map Navigation with Mobile Devices: Virtual versus Physical Movement with and without Visual Context Categories and Subject Descriptors," In ICMI' 07, pp. 146–153.

[25] Rohs, M., Schöning, J., Schleicher, R., Essl, G., Naumann, A., and Krüger, A., "Impact of Item Density on Magic Lens Interactions," In MobileHCI '09, pp. 1-4.

[26] Wagner, D., Reitmayr, G., Mulloni, A., Drummond, T., and Schmalstieg, D., "Pose tracking from natural features on mobile phones," In ISMAR '08, pp. 125– 134.